Comparative Effectiveness Review
Number 158

Effective Health Care Program

Diagnosis of Gout

Executive Summary

Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at **www.effectivehealthcare.ahrq.gov/reports/final.cfm**.

Background

Condition

Gout is a form of inflammatory arthritis characterized by acute intermittent episodes of synovitis presenting with joint swelling and pain; the episodes are referred to as acute gouty arthritis flares or attacks. The condition may progress to a chronic and persistent condition, with development of tophi (solid deposits of monosodium urate [MSU] crystals in joints, cartilage, tendons, bursae, bone, and soft tissue), a condition called chronic tophaceous gout. There is no clear distinction between acute intermittent and chronic intermittent conditions, whereas the advanced stage of gout is characterized by more persistent joint manifestations and tophi (either clinically evident or hidden within the joint).

Gout is the most common form of inflammatory arthritis, and the prevalence has been increasing. The most recent estimate of prevalence among adults in the United States, based on data from the 2007–08 National Health and Nutrition Examination Survey (NHANES), is 3.9 percent (8.3 million individuals), ranging from 2.0 percent in women to 5.9 percent in men,[1] an increase over that of previous NHANES data cycles. The rise in the prevalence of gout has paralleled the increase in prevalence of comorbid conditions associated with hyperuricemia (the primary

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Effective Health Care

risk factor for gout), including obesity, hypertension, hypertriglyceridemia, hypercholesterolemia, type 2 diabetes, metabolic syndrome, chronic kidney disease, and renal insufficiency. Increased use of medications that increase the risk for developing hyperuricemia (e.g., thiazide diuretics, low-dose aspirin, or their combination) may further explain the increasing prevalence of gout.

In a 2013 study that analyzed data from several national surveys administered from 2002 to 2008, the number of ambulatory care visits attributable to gout was estimated to be 7 million visits annually, with 2 million attributable to acute attacks. (The rate more than doubled from 2002 to 2008.) The total annual ambulatory care costs associated with gout (visits and medications) were estimated at $933 million (in 2009 dollars). Drug expenditures accounted for 61 percent of the total costs.[2]

In addition to gout, the types of inflammatory arthritis include rheumatoid arthritis, septic arthritis, inflammatory episodes of osteoarthritis, and calcium pyrophosphate dihydrate crystal deposition disease (CPPD, formerly known as pseudogout). Patients with any of these types of arthritis can present with clinically similar signs and symptoms, but the conditions have different treatments, and incorrect diagnosis can have serious outcomes. For example, missing a case of septic arthritis can lead to joint damage and septic shock. A major challenge for effective gout management, particularly in the primary care and urgent/emergent care setting where most gout patients are managed, is distinguishing gout from these other conditions. Inappropriate or delayed treatment can incur serious complications.

Etiology of Gout

The driving force behind acute episodes of gout is hyperuricemia, defined as an elevated serum uric acid (more accurately referred to as "serum urate" for the salt form that occurs in the serum) concentration greater than 6.8 mg per deciliter in men and greater than 6.0 in women. Hyperuricemia is most commonly the result of inadequate renal excretion of uric acid or, less commonly, uric acid overproduction. (Uric acid is a breakdown product of dietary or endogenous purines.) Hyperuricemia leads to formation and deposition of MSU crystals, which preferentially deposit in joints, tendons, and bursa spaces. For reasons that remain unclear, only a small proportion of individuals with hyperuricemia go on to develop gout. For others, hyperuricemia remains asymptomatic.[3] The prevalence of hyperuricemia ranges from 21.2 percent in men to 21.6 percent in women, 4 to 10 times as high as the prevalence of gout.[4]

The causes of gout are multifactorial, including a combination of genetic, hormonal, metabolic, pharmacologic, comorbid (renal disease), and dietary factors. Family history, advancing age, male sex, or, in women, early menopause have been associated with a higher risk of gout and/or gout flares.[5] Dietary risk factors for gout include consumption of purine-rich foods or drinks, including alcohol, meat, and seafood, and consumption of sugar-sweetened soft drinks and foods high in fructose. Dairy foods and coffee have been associated with a lower risk of incident gout and in some cases a lower rate of gout flares. However, the role of diet in the etiology and treatment of gout is a topic of considerable research and will be reviewed in a separate systematic review.

Diagnostic Strategies

The majority of individuals with gout are initially seen, diagnosed, and treated in primary and urgent care settings. Thus primary care physicians (PCPs) and emergency medicine physicians are the most likely practitioners to see patients with symptoms suggestive of an acute attack of gout but with no prior diagnosis. Such patients may be experiencing a first attack (early-stage gout) or may have experienced numerous attacks and have more advanced gout.

Some researchers have argued the need for laboratory assessment of synovial (joint) fluid MSU crystals in the presence of an acute inflammatory arthritis for a definitive diagnosis of gout, and MSU crystal analysis has been regarded as the gold standard against which other potential diagnostic methods are measured. However, joint aspiration can be technically difficult to perform and painful to the patient, and is often deferred in primary and urgent care settings, to be conducted by a specialist (e.g., a rheumatologist or orthopedic surgeon).[6] In addition, the accuracy of synovial fluid analysis may be affected by a number of factors (patient, practitioner, and analyst related).[7,8] A 2009 study found that unguided needle insertion in the toe is often inaccurate.[9] At least three studies have found wide variation in the accuracy of assessment of synovial fluid crystals (both MSU and calcium pyrophosphate) and white blood cells across hospital laboratories,[10-12] which could potentially be caused by patient differences, differences in skill levels of the practitioners drawing or analyzing the samples, or differences in sample handling. A 1999 systematic review

on the accuracy of MSU crystal analysis in synovial fluid[13] concluded that MSU analysis had poor sensitivity, specificity, and reproducibility. A 2013 systematic review of the accuracy of methods for detecting MSU in synovial fluid concluded that storage of samples at room temperature resulted in a decrease in MSU concentration over time compared with refrigeration[14] but could not draw any conclusions about the role of personnel. Evidence from a 2011 survey of rheumatologists suggests that synovial fluid analysis is underused in the rheumatology setting as well.[15]

Instead of analyzing MSU crystals in synovial fluid, PCPs and emergency medicine physicians tend to rely on clinical algorithms comprising some combination of clinical signs and symptoms to diagnose an acute episode of gout. These clinical signs and symptoms include rapid development of inflammation and pain, erythema, monoarthritis, response to administration of the drug colchicine, and symptoms in the first metatarsophalangeal joint, among others (with synovial fluid culture sometimes used to rule out septic arthritis and other potential causes for inflammatory arthritis).

Attempts to standardize and validate such clinical diagnostic algorithms date back to the 1960s.[16] Most of these algorithms were not developed for diagnostic purposes but for classification of gout. Concurrent with this review, the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) are collaborating to update and evaluate classification criteria for gout. Distinct from diagnostic criteria, classification criteria are intended to ensure the correct identification and staging of patients with a particular disease condition (especially patients in the early stages of the disease) for the purpose of enrollment in studies of disease management.[17]

Therefore, a question of importance is whether any combination of clinical signs and symptoms and laboratory tests accessible in the primary or acute care setting (which we refer to as a "clinical algorithm" or "clinical diagnostic algorithm") will have good predictive value compared with tests such as joint aspiration and synovial fluid analysis for MSU, both to correctly diagnose gout and to rule out other causes of joint inflammation, particularly septic arthritis and calcium pyrophosphate deposition disease, for patients presenting with an acute episode of inflammatory arthritis.

Imaging modalities have also been assessed for both diagnosis and classification of gout. These techniques include plain radiographs and newer techniques such as ultrasound and dual-energy computed tomography (DECT), which are just beginning to be used to diagnose gout in some settings.[18] Therefore, another question of importance for gout diagnosis is how these newer methods compare with joint aspiration and synovial fluid MSU analysis in their predictive value for the initial diagnosis of gout and whether they provide any additive value over the use of MSU analysis or clinical signs and symptoms alone.

The safety of tests used to diagnose gout also needs to be considered. Potential safety concerns include acute physical discomfort from joint aspiration and long-term effects (e.g., from accumulated radiation exposure). Other concerns are the potential effects of misdiagnosis. These effects could include delay in initiating or failure to initiate appropriate treatment for gout, delay in initiating treatment for the actual disorder if it is not gout, or incorrect initiation of treatment for another disorder (e.g., hospitalization and administration of intravenous antibiotics for suspected joint sepsis) when the patient has gout.

Therefore, we have undertaken a systematic review of studies examining the accuracy and safety of tests used to diagnose gout—including algorithms combining physical signs and symptoms, serum urate, ultrasound, plain radiography, and DECT—compared with synovial fluid MSU analysis. The primary focus of this review is on tests that can be used in the primary care or urgent/emergent care setting for an initial diagnosis of gout.

The aim of this review is to help inform clinical decisionmaking for patients and providers and to improve the quality of care for patients who present with previously undiagnosed gout in the primary and acute care setting.

Scope and Key Questions

Scope of the Review

The purpose of this review is to assess the evidence on the validity and safety of tests for diagnosing gout—including clinical signs and symptoms (individually and in combination as a clinical diagnostic algorithm), DECT, ultrasound, and other imaging methods—compared with aspiration of synovial fluid from involved joints and analysis of MSU crystals using polarized light microscopy. Because concerns have been raised about the accuracy of MSU crystal analysis itself, the review also assesses the evidence that practitioner type may affect the outcomes of MSU analysis. The Agency for Healthcare Research and Quality (AHRQ) assigned this report to the Southern California Evidence-based Practice Center (Contract No. 290-2012-00006-I). A protocol for the review was

posted on the AHRQ Web site on July 17, 2014, at www://effectivehealthcare.ahrq.gov/ehc/products/564/1937/gout-protocol-140716.pdf. The protocol was approved by the AHRQ Center for Evidence and Practice Improvement.

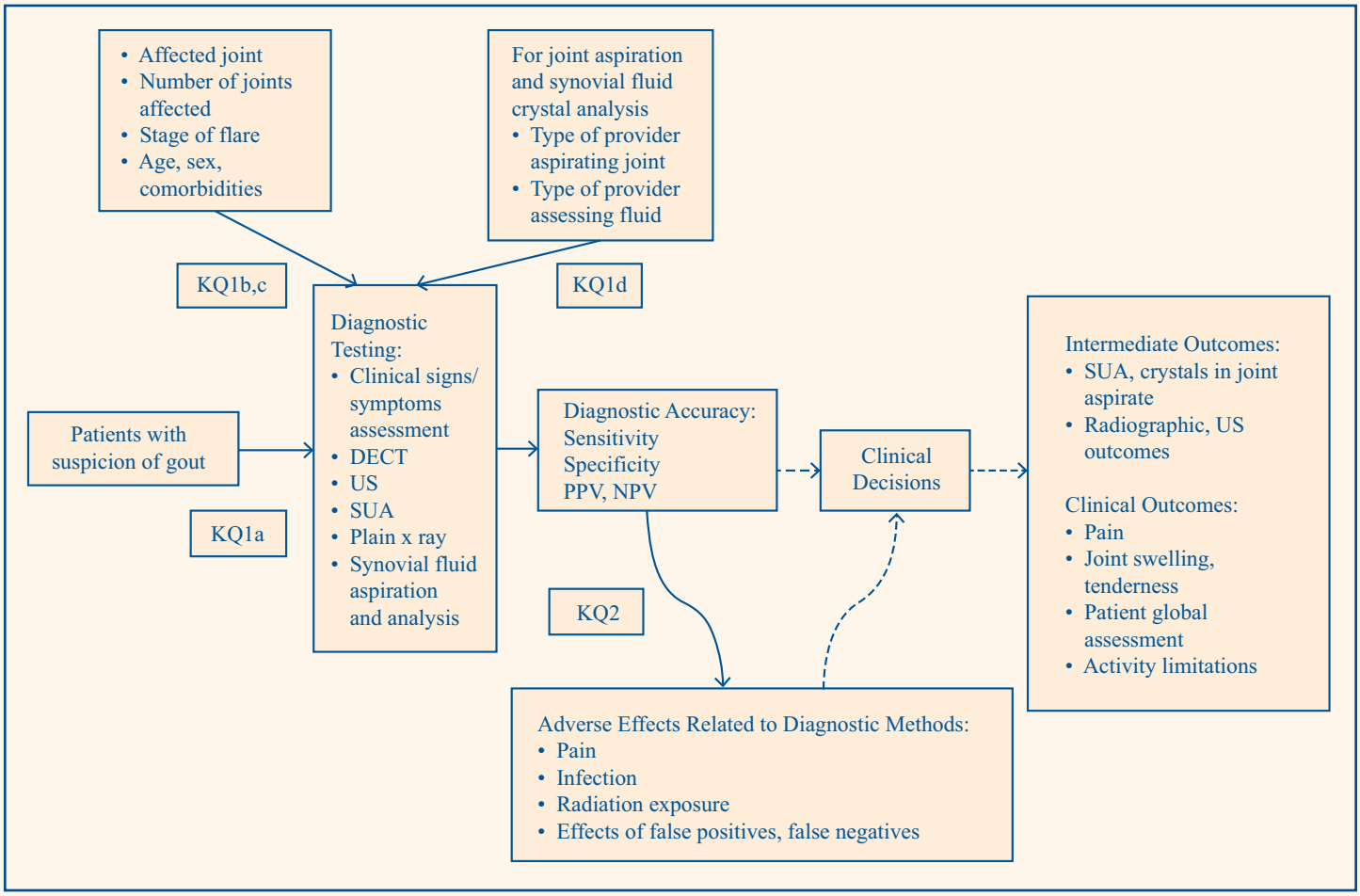
Key Questions

Figure A shows an analytic framework to illustrate the population, interventions, outcomes, and adverse effects that guided the literature search and synthesis for this project. The framework shows the population of interest, patients with symptoms suggestive of possible gout, undergoing any of a number of potential diagnostic tests, whose validity is the subject of Key Question 1a. Patient-level factors that might affect the accuracy of these tests are the topics of Key Questions 1b and 1c. Provider factors that might affect the accuracy of one specific test, MSU analysis, are the topic of Key Question 1d. Key Question 2 assesses potential adverse effects that might be associated with testing: short- and long-term harms from the test procedures themselves, and outcomes associated with misdiagnosis. The dotted lines indicate possible outcomes; for example, diagnostic accuracy and adverse effects of testing might affect clinical decisionmaking, which might in turn affect intermediate and clinical outcomes.

Key Question 1.

a. **What is the accuracy of clinical signs and symptoms and other diagnostic tests (such as serum uric acid, ultrasound, computed tomography (CT) scan, DECT, and plain x ray), alone or in combination, compared with[a] synovial fluid analysis in the diagnosis of acute gouty arthritis, and how does the accuracy affect clinical decisionmaking, clinical outcomes and complications, and patient-centered outcomes?**

Figure A. Analytic framework



DECT = dual-energy computed tomography; KQ = Key Question; NPV = negative predictive value; PPV = positive predictive value; SUA = serum uric acid

[a] Using monosodium urate crystal analysis of synovial fluid as the reference standard.

b. How does the diagnostic accuracy of clinical signs and symptoms and other tests vary by affected joint site and number of joints?

c. Does the accuracy of diagnostic tests for gout vary by duration of symptoms (i.e., time from the beginning of a flare)?

d. Does the accuracy of synovial fluid aspiration and crystal analysis differ by (i) the type of practitioner who is performing the aspiration and (ii) the type of practitioner who is performing the crystal analysis?

Key Question 2. What are the adverse effects (including pain, infection at the aspiration site, radiation exposure) or harms (related to false positives, false negatives, indeterminate results) associated with tests used to diagnose gout?

Methods

Criteria for Inclusion/Exclusion of Studies in the Review

This report is based on a systematic search for prospective or cross-sectional studies that compared the sensitivity and specificity of tests used to diagnose gout, preferably against joint aspiration and synovial fluid assessment for MSU crystals, in populations of adults 18 years of age or older suspected of having gout but not previously diagnosed. (See Table A in the Results section.) We also included studies that assessed patient and practitioner factors that affect the diagnostic accuracy of these tests or assessed harms associated with the tests, and studies that examined particular factors that potentially affect the sensitivity or specificity of tests (joints involved, duration of symptoms).

Tests of interest included algorithms comprising clinical or laboratory examination for physical signs, symptoms, and history; serum uric acid; US; DECT; and plain radiography. The comparator of primary interest was synovial fluid analysis of MSU crystals using polarized light microscopy. However, if no such studies could be identified for a diagnostic test of interest, studies were also included if some or all of the participants were diagnosed using the ACR criteria for gout diagnosis and classification or another validated set of diagnostic or classification criteria as a reference standard (comparator).

Studies were excluded if participants had already been definitively diagnosed with gout prior to enrollment (to ensure that the patient populations were as similar as possible to patients who would be seen in the primary or urgent/emergent care setting), or if the comparator was individual physician opinion or was not identified. Inclusion criteria are further described in terms of PICOTs (populations, interventions, comparators, outcomes, timing, and settings), a framework used in systematic reviews to categorize inclusion and exclusion criteria).

Outcomes of interest were the comparative accuracy of the test results (as measured by the sensitivity and specificity or the positive and negative predictive value of the test in question), intermediate outcomes such as lab and radiographic test results, clinical decisionmaking that resulted from a diagnosis, short-term clinical (patient-centered) outcomes such as a change in pain and joint swelling that resulted from a diagnosis, and any adverse events (including adverse patient experiences such as pain or infection at the aspiration site, effects of radiation exposure, and the results of a false-positive or false-negative diagnosis) associated with the test. Prospective cohort, cross-sectional, and case-control (if needed) studies were included to address Key Question 1 (accuracy of test and factors that affect accuracy). Prospective cohort, cross-sectional, and case-control studies, as well as case series of any size and case reports of rare adverse events, were included if they addressed Key Question 2 (adverse events or other negative outcomes in individuals undergoing testing).

The PICOTS for studies included in this review are as follows.

Population(s) (Key Questions 1 and 2):

- Adults (18 years and over) presenting with symptoms (e.g., an acute episode of joint inflammation) suggestive of gout but without a prior gout diagnosis, including the following subgroups:
  - Male and female patients
  - Patients with longer versus shorter duration of symptoms
  - Patients with comorbidities, including hypertension, type 2 diabetes, and kidney disease (renal insufficiency)
  - Patients with osteoarthritis, septic arthritis, calcium pyrophosphate deposition disease, or previous joint trauma
  - Individuals with a family history of gout

Interventions (index tests) (Key Questions 1 and 2):

- Clinical history and physical exam
- Serum urate assessment
- US
- DECT
- Plain x ray

- Joint aspiration by physicians and synovial fluid analysis using polarizing microscopy (by physicians or laboratory personnel)
- Combinations of these tests as identified in the literature

Comparators (reference tests):

- Joint synovial fluid aspiration and microscopic assessment for MSU crystals (Key Questions 1a–c and 2)
- Joint synovial fluid aspiration and microscopic assessment for MSU crystals performed by a practitioner with a different level of expertise or experience, such as rheumatologist, laboratory personnel (Key Question 1d)

Outcomes:

- Diagnostic accuracy of clinical signs and symptoms, ultrasound, DECT, and plain radiographs compared with joint aspiration and synovial fluid analysis (Key Question 1)
  - Sensitivity/specificity, true positives/true negatives, area under the curve
  - Positive and negative predictive value, positive/negative likelihood ratios
- Clinical decisionmaking (Key Question 1)
  - Additional testing
  - Pharmacologic/dietary management
- Intermediate outcomes (Key Question 1)
  - Serum urate
  - Synovial fluid crystals
  - Radiographic or ultrasound changes
- Clinical outcomes (Key Question 1)
  - Pain, joint swelling, and tenderness
  - Patient global assessment and activity limitations (Key Questions 1 and 2)
- Adverse effects of the tests, including—
  - Pain, infection, and radiation exposure
  - Effects of false positives or false negatives (Key Question 2)

Timing:

- For clinical outcomes of symptom relief: 1–2 days minimum (Key Question 1)
- Early in an attack versus later or post-attack (Key Question 1c)
- For adverse events: immediate

Settings:

- Primary care (outpatient) or acute care settings preferred
- Outpatient rheumatology practices/academic medical centers also accepted

Literature Search Strategies for Identification of Studies Relevant to Key Questions

The search strategy was designed by the Southern California Evidence-based Practice Center (EPC) reference librarian in collaboration with our local content expert, who has participated in two systematic reviews on gout;[19,20] it appears in Appendix A of the full report. As recommended by the AHRQ "Methods Guide for Medical Test Reviews,"[21] the searches were conducted without filters specific for diagnostic tests; instead, we used the term "gout" combined with the terms for the diagnostic tests.

We searched PubMed® (January 1, 1946, to November 7, 2014), Embase® (January 1, 1972, to November 7, 2014), the Cochrane Library (January 1, 1945, to November 7, 2014, for the Cochrane Central Registry of Controlled Trials and January 1, 1996, to November 7, 2014, for the Cochrane Database of Systematic Reviews), and the Web of Science™ (January 1, 1980, to November 7, 2014); these dates were selected to replicate the searches conducted as the basis for the 2006 EULAR Guidelines on Diagnosis and Management of Gout.[22] We also included any relevant studies identified in the searches we conducted for a simultaneous review on management of gout if they were not already identified in the searches for this review. Finally, we asked the Technical Expert Panel (TEP) to assess our list of included studies and to provide references for any studies they believed should also be included.

We searched ClinicalTrials.gov and the Web of Science for recently completed studies and unpublished or non–peer-reviewed study findings. Searches were not limited by language of publication: non–English-language studies that met the inclusion/exclusion criteria based on a review of an English-language abstract were screened further in full text if translators could be identified with reasonable effort. We also contacted manufacturers of diagnostic equipment (polarizing microscopes, sonography equipment, DECT, and serum uric acid test kits) for unpublished data specific to the use of their equipment or tests for gout diagnosis.

An update search was conducted on November 7, 2014, after submission of the draft report for peer review. We transferred the output of the literature searches to DistillerSR™ for screening. Article titles and abstracts identified by the searches were independently screened by two literature reviewers using the predetermined inclusion and exclusion criteria, and those selected by either reviewer were accepted without reconciliation for further full-text review.

Two reviewers independently conducted full-text review to exclude articles that provided no usable data, reported the same data as another article, or enrolled participants with established gout diagnoses. Disagreements regarding inclusion at the full-text stage were reconciled with the input of the project lead when necessary.

We identified a small number of relatively recent systematic reviews on various aspects of gout diagnosis. In most cases, we used these reviews to identify references we had missed; however, if the review was of high quality, addressed a subquestion of interest, and included all the literature on the topic, we included it as a data source after assessing its quality. We also searched the reference lists of included studies for additional titles that appeared to meet our inclusion criteria and screened these articles for inclusion. For studies of apparent interest reported in meeting abstracts (conference proceedings), we searched for peer-reviewed publications of the findings. If findings had not yet been published in a peer-reviewed journal, we reserved them and cited them in the Discussion in suggestions for future research.

Data Abstraction and Data Management

Two reviewers independently abstracted study-level details from articles accepted for inclusion in DistillerSR, and any disagreements were reconciled with the input of the project leader, Southern California EPC director, or local subject-matter expert, if needed. Studies provided by manufacturers or suggested by peer reviewers underwent the same process, as did studies identified in update searches.

Assessment of Methodological Quality of Individual Studies

The risk of bias (study quality) of individual included studies was assessed independently by two reviewers using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 tool,[23,24] and assessments were reconciled, with any disagreements mediated by the project lead. We used AMSTAR (A Measurement Tool to Assess Systematic Reviews) to assess the quality of existing systematic reviews that we included;[25] AMSTAR assessments were also conducted independently by two reviewers and reconciled.

Data Synthesis/Analysis

For studies that assessed ultrasound, DECT, or another radiographic method, we extracted and reported sensitivity, specificity, positive and negative predictive value, and area under the curve/receiver-operating characteristics, if reported.

Studies were considered for meta-analysis if the number of true positives, true negatives, false positives, and false negatives was reported or could be calculated; studies were similar enough with respect to outcome measures, participants, and tests; and they assessed the validity of an alternative diagnostic method against that of analysis of MSU crystals in synovial fluid. The number of studies we identified precluded pooling; therefore, outcomes are described narratively in the full report, stratified by test comparisons of interest and study design. All included studies are also described in summary tables in the full report.

Grading the Strength of the Body of Evidence for Each Key Question

We assessed the overall strength of evidence for each conclusion using guidance suggested by AHRQ for its Effective Health Care Program.[26] This method is based on a method developed by the GRADE (Grading of Recommendations, Assessment, Development and Evaluations) Working Group. The evidence grade is usually based on five required domains:

- Study limitations were assessed based on the risk-of-bias assessments for all studies that contribute to a conclusion.
- Consistency was determined by comparing the relative sensitivities and specificities because we did not pool studies.
- Directness is a measure of whether the evidence being assessed reflects a single direct link between the interventions of interest and the ultimate health outcome under consideration.
- Precision, a measure of the confidence intervals in a pooled analysis, also was not assessed in this review.
- Publication bias was assessed only for studies for which data were pooled.

Based on the domains we included, we classified the strength (grade) of evidence as follows:

- High = Further research is unlikely to change our confidence in the estimate of effect.
- Moderate = Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
- Low = Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
- Very low/insufficient = Any estimate of effect is very uncertain.

Applicability

Applicability is a measure of the extent to which the participants, interventions, and outcome measures are similar to those of the population of interest and care settings for which the outcomes are intended. We assessed applicability based on the inclusion and exclusion criteria described in the PICOTS, which included the study population age, sex, health profiles (including comorbidities as well as duration of symptoms and number of affected joints, when relevant), tests, gold standards, study settings, and provider types.[27] Thus we would assign higher priority to studies of adult populations being seen in primary/urgent/emergent care settings for first or subsequent episodes of symptoms suggestive of gout than to studies of patients in an academic rheumatology department.

Peer Review and Public Commentary

A draft version of the report was posted for peer review on November 4, 2014, and revised in response to reviewer comments.

Results

This section first describes the results of the literature searches, followed by descriptions of the studies that met inclusion criteria for each of the Key Questions and the key points (conclusions).

Results of Literature Searches

Our searches identified 3,646 titles/abstracts, of which 3,391 were excluded for the following reasons: participants not human (129); diagnostic methods beyond the scope of the review (129); not gout diagnosis or management (1,801); no original data or nonsystematic reviews (374); conference proceedings, presentations, or abstracts (11); case reports with sample sizes of fewer than 10 (415); population under age 18 (5); renal transplant or end-stage renal disease patients (12); titles with no abstracts (based on a survey of a random sample of 10% of these titles, for which full-text articles or reports were obtained and all were rejected as letters, commentaries, or nonsystematic reviews with no original data) (252); and gout management only (263). (See the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] diagram, Figure B.)

We reviewed 255 full-text articles, of which 234 were excluded for the following reasons: participants not human (2); diagnostic methods beyond the scope of the review (44); not gout diagnosis or management (69); no original data (29); conference proceedings, presentations, or abstracts not identified as such by title and abstract review (38); case reports with sample size fewer than 10 (17); gout management only (13); no reference standard reported or not all patients received the reference standard (7). We were unable to obtain articles for 15 studies.

Our search of ClinicalTrials.gov for gout-related research identified 152 entries, none of which were relevant to this review.

None of the manufacturers of imaging equipment or laboratory test kits used in the diagnosis of gout who were contacted for information responded to requests. A notice placed in the *Federal Register* requesting such information also received no responses.

We include the results of 17 original studies[16,18,28-42] and 4 systematic reviews[43-46] in our evidence synthesis. Seventeen studies answer Key Question 1, and two studies answer Key Question 2. Results are shown by Key Question.

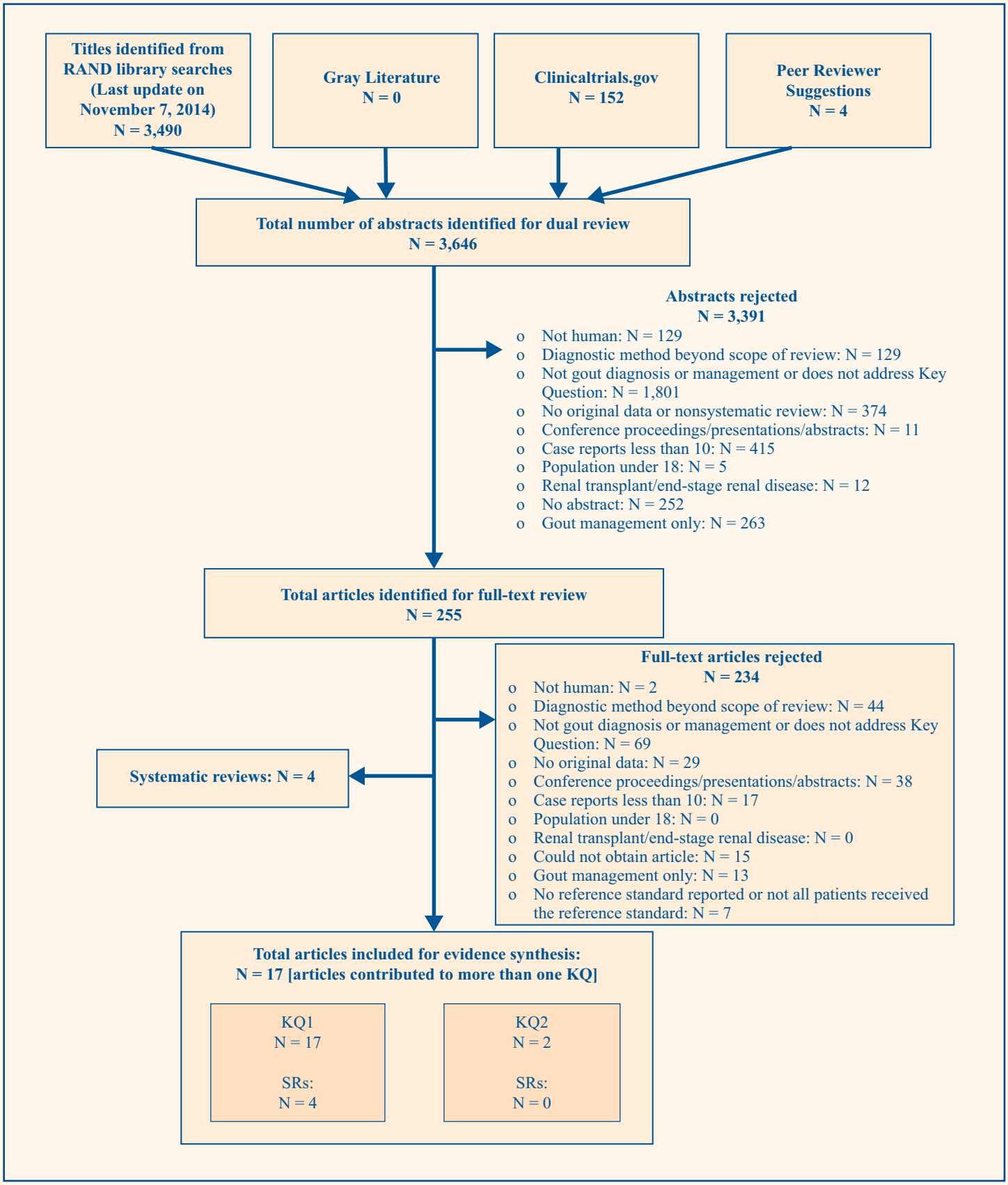
The findings of the review are summarized below and in Tables B and C.

Key Question 1.

a. What is the accuracy of clinical signs and symptoms and other diagnostic tests (such as serum uric acid, ultrasound, CT scan, DECT, and plain x ray), alone or in combination, compared with synovial fluid analysis, in the diagnosis of acute gouty arthritis, and how does the accuracy affect clinical decisionmaking, clinical outcomes and complications, and patient-centered outcomes?

b. How does the diagnostic accuracy of clinical signs and symptoms and other tests vary by affected joint site and number of joints?

Figure B. Literature flow diagram



KQ = Key Question; SR = systematic review

c. Does the accuracy of diagnostic tests for gout vary by duration of symptoms (i.e., time from the beginning of a flare)?

Description of Included Studies

We identified 15 original studies that met our inclusion criteria for studies on the comparative effectiveness of methods for the diagnosis of gout: 9 studies assessed the sensitivity and specificity of combinations of clinical signs and symptoms (clinical algorithms),[16,31-34,39-42] 3 assessed the use of DECT,[18,28,30] and 4 assessed the use of ultrasound (1 study compared ultrasound and DECT).[30,35,36,38] We also identified four prior systematic reviews: one that addressed a clinical algorithm,[46] two that assessed the use of imaging for diagnosis of gout,[43,45] and one on sex differences in gout diagnosis.[44]

The nine studies that assessed the use of clinical algorithms compared the predictions based on six clinical algorithms (Table A) with assessment of synovial fluid MSU crystals in all or most enrolled patients, or at least in those believed to have gout. (In the latter case, patients who were considered not to have gout had to have another condition confirmed by a validated diagnostic criterion.) These studies, which dated from 1977 or later, enrolled from 82 to 983 adult patients, both male and female. All studies were conducted in academic rheumatology departments, although several of the studies purposely enrolled patients who were referred by PCPs.

The three studies that assessed the use of DECT compared the predictions based on these imaging studies with assessment of synovial fluid MSU crystals, with a validated clinical algorithm, or with some combination of the two reference standards. These studies dated from 2011 to 2014 and enrolled from 31 to 94 patients with suspected gout. All studies were conducted in academic rheumatology departments.

The four studies that assessed the use of ultrasound compared the predictions based on ultrasound signs with assessment of synovial fluid MSU crystals, with a validated clinical algorithm, or some combination. The studies dated from 2008 to 2014 and enrolled from 54 to 105 patients with suspected gout.

Key Points

The key points for Key Questions 1a–c are as follows:

- Few studies that assessed the accuracy of diagnostic clinical algorithms consistently applied the same reference standard (either analysis of MSU crystals in synovial fluid or a single clinical algorithm) to all participants with suspected gout.
- Studies that assessed the use of diagnostic clinical algorithms compared with synovial fluid analysis for MSU crystals reported widely varying sensitivities and specificities. However, two recently developed algorithms (the Diagnostic Rule and the Clinical Gout Diagnosis), the former developed from clinical signs and symptoms used by primary care physicians, reported sensitivities of 88 percent and 97 percent, respectively, and specificities of 75 percent and 96 percent, respectively. The strength of evidence for this conclusion is low; it is based on the identification of three studies that assessed one of the clinical algorithms and two studies that assessed the other one, all in single clinics.
- In three studies that enrolled only patients not previously diagnosed with gout, the sensitivities and specificities of DECT for predicting gout ranged from 85 percent to 100 percent compared with synovial fluid analysis for MSU crystals and from 83 percent to 92 percent compared with a validated clinical algorithm. The strength of evidence for this conclusion is low.
- Ultrasound was more variable than DECT in its ability to detect gout. Four studies of ultrasound showed sensitivities ranging from 37 percent to 100 percent and specificities ranging from 68 percent to 97 percent, depending on the signs assessed and probably related to the duration of the disease. The strength of evidence for this conclusion is low.
- No studies were identified that assessed the validity of serum urate, CT scan, or plain x ray for diagnosing gout. The strength of evidence for these tests is insufficient.
- No studies were identified that directly assessed the effect of joint site or number of affected joints on diagnostic accuracy, although several studies indirectly addressed this question for imaging techniques. The strength of evidence for this question is insufficient for all diagnostic methods.
- No studies were identified that directly assessed the effect of duration of symptoms on the accuracy of diagnostic tests. The strength of evidence for this question is insufficient for all diagnostic methods.

Key Question 1d. Does the accuracy of synovial fluid aspiration and crystal analysis differ by (i) the type of practitioner who is performing the aspiration and (ii) the type of practitioner who is performing the crystal analysis?

Table A. Comparison of components among clinical algorithms for diagnosis of gout

Components	Rome, 1963[a,16]	New York, 1966[b,16]	Wallace, 1977 (ARA/ACR)[42]	EULAR, 2006[22]	Janssens' Diagnostic Rule, 2010[31]	CGD, 2010[c,41]	3e Initiative, 2014[d,47]	Richette Survey[e,39]
Clinical characteristics								
>1 attack of acute arthritis		✓	✓		✓	✓		✓
Maximum inflammation developed within 1 day✓	✓ Painful joint swelling, abrupt onset, clearing 1–2 weeks	✓	✓	✓	✓	✓		
Monoarthritis/oligoarthritis attack			✓			✓		
Redness observed over joints			✓	✓	✓	✓		
1st MTP joint painful or swollen			✓	✓	✓	✓ Podagra	✓ Podagra	
Unilateral 1st MTP joint attack		✓	✓					
Unilateral tarsal joint attack			✓			✓		
Abrupt onset and remission in 1–2 weeks initially		✓						
Response to colchicine—major reduction in inflammation within 48 hours		✓					✓	
Pain intensity ≥9/10								✓
Involvement of toes, foot, or ankle								✓
Treatment with corticosteroids								✓
Treatment with NSAIDs								✓
Resolution of pain <15 days after onset		✓	✓					✓
Tophi (proven or suspected)	✓	✓	✓	✓		✓	✓	✓

Table A. Comparison of components among clinical algorithms for diagnosis of gout (continued)

Components	Rome, 1963[a,16]	New York, 1966[b,16]	Wallace, 1977 (ARA/ACR)[42]	EULAR, 2006[22]	Janssens' Diagnostic Rule, 2010[31]	CGD, 2010[c,41]	3e Initiative, 2014[d,47]	Richette Survey[e,39]
Radiographic								
Asymmetric swelling within a joint on radiograph			✓	✓				
Subcortical cysts without erosions on radiograph			✓	✓				
Joint fluid								
Joint fluid culture negative			✓					
MSU crystals in synovial fluid or tissues[f]	✓						✓	
Comorbid or risk factors								
Hyperuricemia	✓		✓	✓	✓	✓		✓
Male sex					✓			✓
Hypertension or ≥1 CVD					✓			✓
Hypertriglyceridemia								✓

3e = Evidence, Expertise, Exchange; ACR = American College of Rheumatology; ARA = American Rheumatology Association; CGD = Clinical Gout Diagnosis; CVD = cardiovascular disease; EULAR = European League Against Rheumatism; MSU = monosodium urate; MTP = metatarsophalangeal; NSAIDs = nonsteroidal anti-inflammatory drugs

Note: Podagra is gout that involves the big toe.

[a] Meets 2 of the criteria.

[b] MSU crystals in joint fluid or tophus or tissue OR meets 2 of the criteria.

[c] ≥4/8 of the criteria checked.

[d] Guideline 1states that MSU is required for a definitive diagnosis but in its absence, clinical criteria such as those checked can be used or characteristic imaging findings may substitute.

[e] Designed to be administered telephonically by nonphysicians to assess prevalence of gout via patient self-report; treatment questions refer to most prominent episode.

[f] Several algorithms specified presence of MSU crystals as definitive in lieu of other signs.

Description of Included Studies

We identified two original studies that addressed this question directly.[29,37] A 2014 study was identified that retrospectively audited medical records of two Korean academic medical centers to assess factors associated with false-negative synovial fluid MSU results; it focused on the personnel performing the analysis and several other factors.[37] A 1989 study compared the accuracy of an experienced rheumatologist, several medical residents, and several technicians in identifying MSU and calcium pyrophosphate crystals suspended in synovial fluid using polarizing microscopy.[29]

Key Points

The key point for Key Question 1d is as follows:

- Agreement among medical and ancillary health personnel examining synovial fluid using polarizing microscopy for detection of MSU crystals appears to be poor, but it is unclear whether the experience and training of analysts are factors. No studies examined the effect of the type of practitioner performing fluid aspiration on the ability to obtain a sample for analysis. Because of the relatively small number of studies identified, the strength of evidence for definitive influential factors is insufficient.

Key Question 2. What are the adverse effects (including pain, infection at the aspiration site, radiation exposure) or harms (related to false positives, false negatives, indeterminate results) associated with tests used to diagnose gout?

Description of Included Studies

One study was identified that assessed adverse effects associated with tests used to diagnose gout.[28] This study reported no adverse events associated with aspiration of synovial fluid for MSU analysis or the use of DECT.

One study examined the outcomes of delayed diagnosis or misdiagnosis of gout in two academic medical centers in South Korea.[37]

Key Points

The key points for Key Question 2 are as follows:

- Potential adverse effects that might be associated with diagnostic tests for gout include pain, infection at the aspiration site, or the short- or long-term effects of radiation exposure. No studies were identified that documented any adverse events associated with diagnostic tests included in this report. The strength of evidence for this conclusion is low, based on one study that reported no adverse events associated with joint fluid aspiration for MSU analysis or DECT, and no studies that reported on adverse events associated with ultrasound or clinical examination.
- Missed diagnosis or delayed diagnosis of acute gout (failure to find MSU crystals in synovial fluid) was reported in a retrospective two-center study to be associated with a longer interval between the onset of attack and joint aspiration. A negative MSU finding was associated with higher risk for undergoing arthroscopic drainage, longer hospital stays, and delays in anti-inflammatory treatment. The strength of evidence for this conclusion is insufficient.

Discussion

Findings in Relation to What Is Already Known

Over the past 25 to 30 years, gout diagnosis has been an area of some controversy. Efforts have been aimed at determining whether the assessment of MSU crystals in synovial fluid aspirated from joints is really the gold standard, validating algorithms comprising various combinations of clinical and laboratory criteria, and validating the use of ultrasound and DECT imaging.

The focus of this report is on evaluating the validity and safety of existing diagnostic methods for use in primary, urgent, and emergency care settings, where the majority of gout patients are first seen and diagnosed. Patients who present in these settings with an inflamed joint and who have not had a prior diagnosis of gout (or another rheumatic condition) are almost certainly having an acute attack, which may be the first or the latest of a number of attacks. Thus, they may be in an early stage of the disease, or at least will be less advanced in the disease process than patients seen in the rheumatology setting. Important considerations in diagnosing gout in these patients include ensuring that criteria are sensitive enough to diagnose less advanced disease and specific enough to rule out other conditions, such as septic arthritis and calcium pyrophosphate deposition disease.

Monosodium Urate Crystal Assessment

The assessment of MSU crystals in synovial fluid for the diagnosis of gout has problems, as noted in the Background section and confirmed by several studies we reviewed, suggesting that it is a suboptimal gold standard against which to measure potential diagnostic methods.[29,37] Further confirming these findings, an abstract presented at the 2013 EULAR meetings on a study that tested the competence of a group of rheumatologists, lab

technicians, and rheumatology residents in identifying MSU and calcium pyrophosphate crystals found that fewer than half identified all samples correctly and that rheumatologist, resident, and technician performance was fairly comparable, although residents performed much more poorly on identification of calcium pyrophosphate crystals.[48]

Nevertheless, recent guidelines continue to recommend the use of MSU assessment for definitive diagnosis. For example, the 2011 Postgraduate Medicine guidelines for diagnosis of gout (which aimed to update the EULAR 2006 guidelines) emphasize that diagnosis based on clinical signs and symptoms alone has reasonable accuracy when patients have typical presentation of gout but that MSU constitutes the definitive diagnosis.[49] (Neither the 2011 Postgraduate Medicine guidelines nor the EULAR 2006 guidelines have been clinically validated.) The 2014 3e (Evidence, Expertise, Exchange) initiative is a multinational effort to promote evidence-based practice. The 3e recommendations on the diagnosis and treatment of gout recognize the use of MSU as the gold standard but also note the difficulty in performing this test under some circumstances, asserting that if MSU cannot be performed, the diagnosis "can be supported by classical clinical features, and/or characteristic imaging findings."[47]

At the 2014 ACR Meeting, new ACR/EULAR diagnostic criteria were presented (updating the 2006 EULAR diagnostic criteria). Based on a systematic review (yet to be published) and consensus panel, the new guidelines advocate the use of MSU for any patient with suspected gout. However, the authors of these latest guidelines also acknowledge the difficulty of assessing MSU and note that, in its absence, a combination of clinical signs and symptoms is suggestive of, but not definitive for, gout.[50]

Accuracy of Algorithms Comprising Clinical Signs and Symptoms for the Diagnosis of Gout

This review identified a series of algorithms, some intended for classification of gout for research purposes (but used in diagnosis as well) and some intended for diagnosis. Comparing the more recent diagnostic algorithms with the earlier algorithms highlights the likely importance of patient population and duration of disease in determining diagnostic criteria. The Diagnostic Rule and the Clinical Gout Diagnosis were developed and validated on patients first identified in primary care; these patients were likely to be in an earlier stage of the disease than the patients on whom earlier diagnostic criteria, such as the ACR criteria, were based. The patients in the earlier validation studies were hand-picked by rheumatologists, which would have increased the sensitivity of the tests compared with their use on a more typical population with a less certain diagnosis.

The incremental utility of MSU over clinical diagnostic criteria alone was recently assessed and compared in patients with shorter (2 years or less) and longer durations of symptoms (history of attacks). This study compared the sensitivities of the classification criteria that include the use of MSU (the Rome, New York, American Rheumatology Association, and Clinical Gout Diagnosis criteria) with and without the MSU findings. They found that, in patients with shorter symptom duration, inclusion of MSU assessment improved sensitivity considerably over the same criteria without MSU. Nevertheless, the sensitivities of the CGD criteria without including an MSU assessment and the Diagnostic Rule (which does not include MSU) were still fairly high (87.2% and 87.9%, respectively). The sensitivities of all clinical diagnostic and classification criteria are greater for patients with symptom duration longer than 2 years than for newer patients. In addition, omission of MSU and reliance on the clinical diagnostic criteria alone resulted in a much smaller decrease in sensitivity for these more advanced patients. None of the studies we identified limited inclusion to patients having a first attack.

Accuracy of DECT for the Diagnosis of Gout

DECT is a noninvasive study method that can detect urate deposits in joints, tendons, bursa, and soft tissues. The radiographic signature of urate can be distinguished from that of calcium. DECT requires special machines and software to process the images and currently is not widely available. Radiation exposure is not greater than standard CT scanning and is limited to extremities, which are not radio-sensitive organs.

Studies assessing the diagnostic utility of DECT are promising, generally demonstrating high sensitivity and specificity for gout. However, we identified only a small number of studies on patients without previous diagnoses of gout.

A recent publication[28] sought to determine the additive value of DECT to a clinically unclear presentation among 30 patients. Of these 30, 14 had a positive DECT, and of those 14, 11 of 12 (2 patients refused aspiration) had crystal confirmation of gout using ultrasound-guided aspiration. In a group of 40 patients seen in the same clinic whose gout was confirmed with MSU assessment, all 4 patients with false-negative DECT had new-onset gout (first attack and symptom duration <6 months). A 2011 study prospectively studied inflammatory monoarthritis

patients, demonstrating high sensitivity and specificity for crystal-confirmed gout cases.[18]

The summary of the literature demonstrates that DECT can be both specific and sensitive for gout. Utility of DECT may be best for evaluating urate burden in established gout patients. Limited data suggest that for patients with recurrent attacks of inflammatory monoarthrities or oligoarthritis for whom the question of gout is unresolved (for example, no fluid available for aspiration or negative study), DECT should demonstrate good diagnostic value. However, for patients with a first inflammatory monoarticular attack (due to gout), DECT may not be sensitive. The lack of availability of DECT machines in most regions also may limit application of this technology.

Accuracy of Ultrasound for the Diagnosis of Gout

Although we identified only a small number of studies assessing the accuracy of ultrasound for diagnosis of gout in patients without a previous diagnosis, its use as a diagnostic test appears to be promising. Sensitivity and specificity for specific ultrasound characteristics or signals (such as the "double contour sign," characteristic intra-articular findings [bright spots or "snow"], and tophaceous findings, or combinations of these signals) were typically high, with one exception. In addition, it is relatively inexpensive, noninvasive, and well accepted by patients.

However, several challenges must be overcome prior to ultrasound being accepted as a standard diagnostic tool. The various signals can present in many different joints, and the analyses we reviewed each used different methodology for identifying which joints they studied. The number of joints studied ranged from a single target (inflamed) joint to 26 joints. Additionally, up to 20 tendon areas and 6 bursae were examined. Such exhaustive scanning is not practical. Some authors[36] described limited systematic evaluation of inflammatory monoarthritis patients with sensitivities and likelihood ratios for specific findings. Nevertheless, even this focused methodology (4 to 6 joints) may be beyond what would be available from most radiology centers, which typically focus on more comprehensive examinations of single joints. The tendency to conduct multisite scans to diagnose and characterize gout appears to be greatest in the rheumatology community.

The low sensitivity reported for the knee double contour sign by Lai and colleagues was attributed to the shorter duration of disease in the included patients,[35] suggesting better diagnostic value in patients with more advanced disease, although another study reported no differences between patients having their first attack and those having had several attacks.[36] Furthermore, we did not find any studies that evaluated the marginal utility of using ultrasound data to diagnose gout above that of using clinical criteria alone or in lieu of joint aspiration.

Thus, the present review confirms the results of several relatively recent systematic reviews on the validity and potential superiority of DECT (and ultrasound) for the diagnosis of gout. However, as the 3e recommendations note, the "availability, cost, and the need for trained personnel and specific equipment" might limit their use in routine clinical practice. Thus, these guidelines seem to suggest that in primary care settings, diagnosis can be based on a set of clinical criteria.[51]

Applicability

Two factors may reduce the applicability of this review.

First, of the studies we identified that assessed the validity of clinical diagnostic algorithms and imaging for the diagnosis of gout, most included at least some participants who had already had a definitive diagnosis. Relatively few studies enrolled only participants with an inflamed joint or even with suspected gout but no established diagnosis. Although the present review excluded studies of individuals with a prior gout diagnosis, we identified no studies that limited inclusion only to patients presenting with a first attack, and few studies considered the duration of the disease or the number of prior attacks in their assessments.

Second, all imaging studies were conducted in a rheumatology setting, usually an academic rheumatology department. Patients seen in this setting may have more advanced disease than those seen in a primary care setting or may have comorbidities that add complexity to their treatment.

Implications for Clinical and Policy Decisionmaking

The findings of this review provide some evidence to support the further development and validation of clinical diagnostic algorithms based on a combination of clinical signs and symptoms for the diagnosis of gout in the primary care setting. The review further supports the use of imaging modalities (ultrasound and DECT) in cases in which a definitive diagnosis cannot be made from signs and symptoms alone.

Limitations of the Comparative Effectiveness Review Process

Assessing the comparative validity of diagnostic tests in systematic reviews presents a number of challenges that are not faced with comparative effectiveness reviews of

treatment strategies. These limitations are magnified by several issues surrounding tests for gout and the natural history of the disease itself. To increase applicability to the specific patient population and health care settings of interest, we limited included studies to those that enrolled previously undiagnosed patients. In doing so, we excluded a number of studies on the use of ultrasound and DECT for monitoring gout or hyperuricemia. Previous systematic reviews on the use of ultrasound and DECT included studies of patients with asymptomatic hyperuricemia and studies of patients with definitive gout diagnoses in various stages of the disease, along with studies of patients with suspected gout but without definitive diagnoses.

Our searches were aimed at identifying studies on gout diagnosis. Searches that identified studies on gout would be expected to identify studies on the differential diagnosis of gout, septic arthritis, calcium pyrophosphate deposition disease, and other such conditions. If a study were aimed at diagnosing patients with a monoarthritis or oligoarthritis, there is a nearly 100-percent chance that the word "gout" would appear, as that would be one possible diagnosis. However, we might have overlooked an occasional study on differential diagnosis of inflammatory joint conditions that was applicable to gout.

In addition, our consideration of unpublished literature was limited. We were unable to obtain information from manufacturers of microscopes and imaging equipment used to diagnose gout. In addition, we did not include conference proceedings as sources of data but cited them in discussing our findings in the context of what is known about gout diagnosis.

Limitations of the Evidence Base

The literature that addresses the diagnosis of gout has numerous limitations that make it difficult to draw firm conclusions. These limitations can be divided into three categories: study volume, design, and reporting quality. We have already addressed some of the issues in the previous discussion. Few studies have attempted to address the diagnosis of gout. Almost no studies have examined the impact of diagnostic test accuracy on decisionmaking (decisions to order further testing or to initiate particular treatments) or any clinical or patient-centered outcomes, and almost no studies addressed adverse events potentially associated with diagnostic testing. Most studies of gout address management issues or monitoring of patients with chronic gout. Of the diagnostic studies we identified, few limited enrollment to patients suspected of having gout or patients with a monoarthritis or some other clinical signs or symptoms that might suggest gout. Many studies enrolled only patients with known gout and included no control group.

Even studies that enrolled patients suspected to have gout or included a control group and employed blinded assessment systematically failed to limit enrollment to patients in their first attack or with recent onset, or did not stratify findings by duration of the condition (as would be ascertained by asking, "How long have you been having these attacks?"). The lack of stratification by duration of condition affects the sensitivity and specificity of both clinical diagnostic algorithms and imaging techniques. Most studies also failed to stratify by other relevant factors, such as time since the onset of the current or most recent flare, sex, and comorbidities. The time since onset of the current flare definitely affects the presence of crystals, as well as clinical signs and symptoms.

No studies tested the validity of combining a clinical diagnostic algorithm comprising clinical signs and symptoms with an imaging test compared with a clinical algorithm or imaging alone. And, as described previously, issues concerning the use of synovial fluid MSU crystal identification as the reference standard abound.

Finally, failure to report important study design details in publications is a further limitation. Studies tended to be vague regarding blinding of assessors and the time lapse between implementation of the index test and reference standard (and the sequence of tests), a critical detail considering the short duration of gout attacks.

Research Gaps

In a 2013 commentary, Dalbeth[17] noted that, thus far, none of the current diagnostic (classification) criteria have been adequately validated: efforts to validate the existing classification criteria have either failed to enroll patients prospectively (i.e., before a definitive diagnosis has been made) or have been limited to very small numbers of patients. The ongoing Study for Updated Gout clAssification cRiteria (SUGAR) project is validating gout classification criteria to improve case ascertainment for recruitment into research studies and for epidemiological purposes. As we suggested in describing limitations of the research base, promising algorithms for diagnosis in the primary care setting, such as the Diagnostic Rule and the Clinical Gout Diagnosis, have limited validation; additional validation is needed in larger, broader populations.

In addition, specific elements of the criteria, such as hyperuricemia, require additional testing. Most clinical diagnostic and classification criteria for gout include

hyperuricemia as a criterion.[16,22,31,39,41,42] However, a 1994 study concluded that serum urate was not a valid criterion for diagnosing gout, as there is no lower level below which gout is not a possibility (and no upper limit beyond which it is a certainty).[52] The 2011 Postgraduate Medicine criteria also excluded hyperuricemia as an element for that reason,[49] and the new 2014 ACR/EULAR criteria include hyperuricemia but state that it should not be the sole criterion on which a diagnosis of gout is made.[50] Thus, further assessment of the effect of hyperuricemia on the sensitivity and specificity of the clinical diagnostic algorithms may be needed.

Patient-level factors that influence test behavior have also been understudied. These include the influence of duration of a flare; number and identity of joints involved; and patient age, sex, and comorbidities. A 2010 systematic review on the diagnosis of gout in women noted that clinical features and risk factors of gout in women differ from those in men.[44] Women have later onset, are more likely to be taking diuretics, have more cardiovascular disease and renal comorbidity, are less likely to drink alcohol, are less likely to have podagra (more involvement of other joints), are more likely to have polyarticular gout, and have less frequent recurrent attacks. These findings suggest the need for different clinical diagnostic criteria for women. Likewise, a number of the clinical diagnostic criteria, including the Diagnostic Rule and the 2014 ACR/EULAR criteria, include cardiovascular comorbidities as a criterion. The sensitivity and specificity of this criterion may need to be established across a broad group of populations.

The findings of Park and colleagues on the effects of gout misdiagnosis[37] suggest that studies are needed on differential diagnosis of gout and other inflammatory joint conditions, particularly septic arthritis and calcium pyrophosphate deposition disease. We identified two recent studies that assessed the validity of a simple laboratory test for the differential diagnosis of gout from septic arthritis. Neither study met our inclusion criteria because the gout diagnosis was made prior to the studies. A 2014 study conducted in Germany analyzed multiple inflammatory markers in serum and synovial fluid drawn from patients seen in a hospital emergency room; gout and septic arthritis were ascertained by synovial fluid aspiration with MSU crystal identification and culture, respectively. Among the markers assayed (e.g., serum uric acid, synovial fluid white blood cells, synovial fluid total protein), synovial fluid lactate had the greatest diagnostic potential to differentiate septic arthritis from gout, followed by glucose and serum uric acid concentrations.[53] A 2014 study conducted in an academic orthopedics department in China found that serum and synovial fluid procalcitonin can both discriminate between septic arthritis and the noninfectious forms of arthritis (gout, rheumatoid arthritis, and osteoarthritis) in the knee, but that synovial fluid procalcitonin is much more sensitive;[54] unfortunately, this assessment would still require joint aspiration. Response to colchicine, which has been suggested as a diagnostic criterion for gout, also does not distinguish gout from other crystal arthopathies. Ultrasound and DECT show some evidence of distinguishing gout from calcium pyrophosphate deposition disease, but further work is needed. Finally, studies are needed that assess the incremental value of ultrasound and DECT imaging over the use of a clinical diagnostic algorithm or even MSU analysis alone. One study assessed the potential additive value of DECT in patients with uncertain diagnosis: the findings suggested that DECT may be a useful adjunct to clinical algorithms among patients with disease of longer duration but not those with new-onset gout (first attack and symptom duration <6 months).[28] Another study purported to assess the added value of ultrasound in a clinical diagnostic algorithm, but this study fell short of actually achieving that outcome.[36] This information will be necessary in determining the importance and the practicality of setting a guideline for referring patients for imaging in making a diagnosis of gout. Of potential utility would be an appropriateness assessment study that creates a panel of possible clinical scenarios of inflammatory joint presentation with the goal of eliciting the most appropriate diagnostic workup for the primary/urgent/emergency care setting.

Conclusions

This review highlights the need for further, broader validation of promising clinical diagnostic algorithms in primary care settings, where the majority of patients with signs and symptoms suggestive of gout, but no definitive gout diagnosis, are likely to be seen. A clinical algorithm with high diagnostic accuracy can ideally form part of a decision tree, with referral of more clinically challenging cases to rheumatologists for more invasive tests or imaging. Research is needed to assess the incremental value of synovial fluid MSU crystal analysis and imaging over that of a diagnostic clinical algorithm. Table B summarizes findings and strength of evidence. Table C summarizes findings on comparative accuracy and safety of gout diagnostic methods.

Table B. Summary of findings and strength of evidence			
Key Question	**Number/Type of Studies**	**Strength of Evidence**	**Findings**
1a. Diagnostic accuracy			
Clinical signs and symptoms (algorithms)	9 observational[16,31-34,39-42]	Low	Tests vary in accuracy compared with synovial fluid aspiration and MSU crystal analysis. Two algorithms based on primary care patients had sensitivities of 88% and 97% and specificities of 75% and 96% but have undergone limited validation.[31,41]
DECT	3 observational 1 systematic review	Low	Sensitivities ranged from 85% to 100% and specificities ranged from 83% to 92% in diagnosing gout.
Ultrasound	4 observational 2 systematic reviews	Low	Sensitivities ranged from 37% to 100% and specificities ranged from 68% to 97%, depending on the ultrasound signs assessed; sensitivity may be lower in patients with early disease.
Other tests	0 studies	Insufficient	None
1b. Influence of number and types of joints involved	0 studies	Insufficient	None
1c. Influence of symptom duration	0 studies	Insufficient	None
1d. Influence of factors on analysis of MSU crystals	2 observational 1 systematic review	Insufficient	Agreement among personnel examining synovial fluid using polarizing microscopy for detection of MSU crystals appears to be poor, but the role of training and experience is unclear. No studies examined the effect of the type of practitioner performing fluid aspiration on the ability to obtain a sample.
2. AEs	2 observational: 1 on AEs associated with 2 diagnostic methods and 1 on implications of misdiagnosis	Low	One study reported that DECT and joint aspiration for MSU analysis were associated with no adverse events.
Implications of misdiagnosis	1 observational study on implications of misdiagnosis	Insufficient	One study reported that missed diagnosis of gout resulted in longer hospital stays, unnecessary surgery, and delayed pharmacological treatment.

AE = adverse event; DECT = dual-energy computed tomography; MSU = monosodium urate

Table C. Summary of findings on comparative accuracy and safety of gout diagnostic methods

Outcomes	Level of Evidence	Findings
Accuracy of Method		
Clinical algorithms based on primary care patients	Low	Sensitivity: 88%–97% Specificity: 75%–96%
US	Low	Sensitivity: 37%–100% Specificity: 68%–97%
DECT	Low	Sensitivity: 85%–100% Specificity: 83%–92%
MSU crystal analysis	NA	Reference standard
Factors Potentially Affecting Accuracy		
Number and/or types of joints involved	Insufficient	No conclusion possible
Patient sex	Insufficient	No conclusion possible
Duration of symptoms (early vs. late disease)	Insufficient	No conclusion possible
Duration of current flare	Insufficient	No conclusion possible
MSU sample handling	Insufficient	No conclusion possible
DECT or US number of views	Insufficient	No conclusion possible
Clinician type–examiner	Insufficient	No conclusion possible
Clinician training or experience	Insufficient	No conclusion possible
Practitioner performing aspiration	Insufficient	No conclusion possible
Facility characteristics	Insufficient	No conclusion possible
Adverse Events		
Adverse events associated with procedures		
US	Insufficient	No conclusion possible
DECT	Low	Evidence suggests that few serious risks are associated with use of DECT for gout diagnosis
MSU crystal analysis	Low	Evidence suggests that few serious risks are associated with use of MSU analysis for gout diagnosis
Adverse events associated with false positives or negatives		
Clinical algorithms	Insufficient	No conclusion possible
US	Insufficient	No conclusion possible
DECT	Insufficient	No conclusion possible
MSU crystal analysis	Insufficient	No conclusion possible

AE = adverse event; DECT = dual-energy computed tomography; NA = not applicable; US = ultrasound
Note: Sensitivity is avoidance of false negatives; specificity is avoidance of false positives.

References

1. Zhu Y, Pandya BJ, Choi HK. Prevalence of gout and hyperuricemia in the US general population: the National Health and Nutrition Examination Survey 2007-2008. Arthritis Rheum. 2011 Oct;63(10):3136-41. PMID: 21800283.
2. Li C, Martin BC, Cummins DF, et al. Ambulatory resource utilization and cost for gout in United States. Am J Pharm Benefits. 2013 March/April;5(2):e46-e54.
3. Doghramji PP, Wortmann RL. Hyperuricemia and gout: new concepts in diagnosis and management. Postgrad Med. 2012 Nov;124(6):98-109. PMID: 23322143.
4. Wallace SL. Hyperuricemia in the diagnosis of gout. J Gen Intern Med. 1989 Mar-Apr;4(2):178-9. PMID: 2651605.
5. Singh JA, Reddy SG, Kundukulam J. Risk factors for gout and prevention: a systematic review of the literature. Curr Opin Rheumatol. 2011 Mar;23(2):192-202. PMID: 21285714.
6. Hamburger M, Baraf HS, Adamson TC 3rd, et al. 2011 Recommendations for the diagnosis and management of gout and hyperuricemia. Postgrad Med. 2011 Nov;123(6 Suppl 1):3-36. PMID: 22156509.
7. Harrold LR, Mazor KM, Negron A, et al. Primary care providers' knowledge, beliefs and treatment practices for gout: results of a physician questionnaire. Rheumatology (Oxford). 2013 Sep;52(9):1623-9. PMID: 23620554.
8. Owens D, Whelan B, McCarthy G. A survey of the management of gout in primary care. Ir Med J. 2008;101(5):147-9. PMID: 18624262.
9. Khosla S, Thiele R, Baumhauer JF. Ultrasound guidance for intra-articular injections of the foot and ankle. Foot Ankle Int. 2009 Sep;30(9):886-90. PMID: 19755074.
10. Hasselbacher P. Variation in synovial fluid analysis by hospital laboratories. Arthritis Rheum. 1987;30(6):637-42. PMID: 3606682.
11. Schumacher HR Jr, Sieck MS, Rothfuss S, et al. Reproducibility of synovial fluid analyses. A study among four laboratories. Arthritis Rheum. 1986 Jun;29(6):770-4. PMID: 3718565.
12. McGill NW, York HF. Reproducibility of synovial fluid examination for crystals. Aust N Z J Med. 1991 Oct;21(5):710-3. PMID: 1759919.
13. Segal JB, Albert D. Diagnosis of crystal-induced arthritis by synovial fluid examination for crystals: lessons from an imperfect test. Arthritis Care Res. 1999 Dec;12(6):376-80. PMID: 11081008.
14. Graf SW, Buchbinder R, Zochling J, et al. The accuracy of methods for urate crystal detection in synovial fluid and the effect of sample handling: a systematic review. Clin Rheumatol. 2013 Feb;32(2):225-32. PMID: 23138881.
15. Perez-Ruiz F, Carmona L, Yebenes MJG, et al. An audit of the variability of diagnosis and management of gout in the rheumatology setting: the Gout Evaluation and Management Study. J Clin Rheumatol. 2011 Oct;17(7):349-55. PMID: 21946457.
16. Malik A, Schumacher HR, Dinnella JE, et al. Clinical diagnostic criteria for gout: comparison with the gold standard of synovial fluid crystal analysis. J Clin Rheumatol. 2009 Feb;15(1):22-4. PMID: 19125136.
17. Dalbeth N, Fransen J, Jansen TL, et al. New classification criteria for gout: a framework for progress. Rheumatology (Oxford). 2013 Oct;52(10):1748-53. PMID: 23611919.
18. Glazebrook KN, Guimaraes LS, Murthy NS, et al. Identification of intraarticular and periarticular uric acid crystals with dual-energy CT: initial evaluation. Radiology. 2011 Nov;261(2):516-24. PMID: 21926378.
19. Khanna D, Fitzgerald JD, Khanna PP, et al. 2012 American College of Rheumatology guidelines for management of gout. Part 1: systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricemia. Arthritis Care Res (Hoboken). 2012;64(10):1431-46. PMID: 23024028.
20. Khanna D, Khanna PP, Fitzgerald JD, et al. 2012 American College of Rheumatology guidelines for management of gout. Part 2: therapy and antiinflammatory prophylaxis of acute gouty arthritis. Arthritis Care Res (Hoboken). 2012;64(10):1447-61. PMID: 23024029.
21. Whiting P, Westwood M, Beynon R, et al. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. J Clin Epidemiol. 2011 Jun;64(6):602-7. PMID: 21075596.
22. Zhang W, Doherty M, Pascual E, et al. EULAR evidence based recommendations for gout. Part I: diagnosis. Report of a task force of the Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis. 2006 Oct;65(10):1301-11. PMID: 16707533.
23. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. PMID: 22007046.
24. Santaguida PL, Riley CM, Matchar DB. Assessing risk of bias as a domain of quality in medical test studies. In: Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EHC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. Chapters available at www.effectivehealthcare.ahrq.gov/reports/final.cfm.
25. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10. PMID: 17302989.
26. Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EHC017. Agency for Healthcare Research and Quality. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. Chapters available at www.effectivehealthcare.ahrq.gov/reports/final.cfm.
27. Atkins D, Chang S, Gartlehner G, et al. Assessing the applicability of studies when comparing medical interventions. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. January 2014. Chapters available at www.effectivehealthcare.ahrq.gov.
28. Bongartz T, Glazebrook KN, Kavros SJ, et al. Dual-energy CT for the diagnosis of gout: an accuracy and diagnostic yield study. Ann Rheum Dis. 2015 Jun;74(6):1072-7. PMID: 24671771.

29. Gordon C, Swan A, Dieppe P. Detection of crystals in synovial fluids by light microscopy: sensitivity and reliability. Ann Rheum Dis. 1989 Sep;48(9):737-42. PMID: 2478085.

30. Huppertz A, Hermann KG, Diekhoff T, et al. Systemic staging for urate crystal deposits with dual-energy CT and ultrasound in patients with suspected gout. Rheumatol Int. 2014 Jun;34(6):763-71. PMID: 24619560.

31. Janssens HJEM, Fransen J, Van De Lisdonk EH, et al. A diagnostic rule for acute gouty arthritis in primary care without joint fluid analysis. Arch Intern Med. 2010;170(13):1120-6. PMID: 20625017.

32. Janssens HJEM, Janssen M, Van De Lisdonk EH, et al. Limited validity of the American College of Rheumatology criteria for classifying patients with gout in primary care. Ann Rheum Dis. 2010;69(6):1255-6. PMID: 19910298.

33. Kienhorst LB, Janssens HJ, Fransen J, et al. The validation of a diagnostic rule for gout without joint fluid analysis: a prospective study. Rheumatology (Oxford). 2015 Apr;54(4):609-14. PMID: 25231179.

34. Kienhorst LB, Janssens HJ, Fransen J, et al. Arthritis of the first metatarsophalangeal joint is not always gout: a prospective cohort study in primary care patients. Joint Bone Spine. 2014 Jul;81(4):342-6. PMID: 24468668.

35. Lai KL, Chiu YM. Role of ultrasonography in diagnosing gouty arthritis. J Med Ultrasound. 2011 March;19(1):7-13. http://dx.doi.org/10.1016/j.jmu.2011.01.003.

36. Lamers-Karnebeek FB, Van Riel PL, Jansen TL. Additive value for ultrasonographic signal in a screening algorithm for patients presenting with acute mono-/oligoarthritis in whom gout is suspected. Clin Rheumatol. 2014 Apr;33(4):555-9. PMID: 24510062.

37. Park JW, Ko DJ, Yoo JJ, et al. Clinical factors and treatment outcomes associated with failure in the detection of urate crystal in patients with acute gouty arthritis. Korean J Intern Med. 2014;29(3):361-9. PMID: 24851071.

38. Rettenbacher T, Ennemoser S, Weirich H, et al. Diagnostic imaging of gout: comparison of high-resolution US versus conventional X-ray. Eur Radiol. 2008 Mar;18(3):621-30. PMID: 17994238.

39. Richette P, Clerson P, Bouee S, et al. Identification of patients with gout: elaboration of a questionnaire for epidemiological studies. Ann Rheum Dis. 2014 May 5. PMID: 24796335.

40. Taylor WJ, Fransen J, Dalbeth N, et al. Performance of classification criteria for gout in early and established disease. Ann Rheum Dis. 2014 Oct 28. PMID: 25351521.

41. Vazquez-Mellado J, Hernandez-Cuevas CB, Alvarez-Hernandez E, et al. The diagnostic value of the proposal for clinical gout diagnosis (CGD). Clin Rheumatol. 2012 Mar;31(3):429-34. PMID: 21979446.

42. Wallace SL, Robinson H, Masi AT, et al. Preliminary criteria for the classification of the acute arthritis of primary gout. Arthritis Rheum. 1977 Apr;20(3):895-900. PMID: 856219.

43. Chowalloor PV, Keen HI. A systematic review of ultrasonography in gout and asymptomatic hyperuricaemia. Ann Rheum Dis. 2013 May;72(5):638-45. PMID: 23291387.

44. Dirken-Heukensfeldt K, Teunissen TAM, van de Lisdonk EH, et al. "Clinical features of women with gout arthritis." A systematic review. Clinical Rheumatology. 2010 Jun;29(6):575-82. PMID: 20084441.

45. Ogdie A, Taylor W, Weatherall M, et al. Imaging modalities for the classification of gout: systematic literature review and meta-analysis. Ann Rheum Dis. 2014 Jun 20. [Epub ahead of print.] PMID: 24915980.

46. Sivera F, Andres M, Falzon L, et al. Diagnostic value of clinical, laboratory, and imaging findings in patients with a clinical suspicion of gout: a systematic literature review. J Rheumatol Suppl. 2014 Sep;92:3-8. PMID: 25180122.

47. Sivera F, Andres M, Carmona L, et al. Multinational evidence-based recommendations for the diagnosis and management of gout: integrating systematic literature review and expert opinion of a broad panel of rheumatologists in the 3e initiative. Ann Rheum Dis. 2014 Feb;73(2):328-35. PMID: 23868909.

48. Berendsen D, Jansen TL, Taylor W, et al. A critical appraisal of the competence of crystal identification by rheumatologists. Ann Rheum Dis. 2014;72(Suppl 3):A981-2.

49. Hamburger MI, Tesser JRP, Skosey JL, et al. Patterns of gout treatment and related outcomes in US community rheumatology practices: the relation between gout flares, time in treatment, serum uric acid level and urate lowering therapy. Arthritis and Rheumatism. 2012;64(Suppl 10):1904.

50. Richette P, Pascual E, Doherty M, et al. Updated EULAR evidence-based recommendations for the diagnosis of gout. Ann Rheum Dis. 2014;(Suppl 2):783-84.

51. Dalbeth N. Management of gout in primary care: challenges and potential solutions. Rheumatology (Oxford). 2013 Sep;52(9):1549-50. PMID: 23771950.

52. Rigby AS, Wood PH. Serum uric acid levels and gout: what does this herald for the population? Clin Exp Rheumatol. 1994 Jul-Aug;12(4):395-400. PMID: 7955603.

53. Lenski M, Scherer MA. Analysis of synovial inflammatory markers to differ infectious from gouty arthritis. Clin Biochem. 2014 Jan;47(1-2):49-55. PMID: 24177196.

54. Wang C, Zhong D, Liao Q, et al. Procalcitonin levels in fresh serum and fresh synovial fluid for the differential diagnosis of knee septic arthritis from rheumatoid arthritis, osteoarthritis and gouty arthritis. Exp Ther Med. 2014 Oct;8(4):1075-80. PMID: 25187799.

Full Report

This executive summary is part of the following document: Newberry SJ, FitzGerald J, Maglione MA, O'Hanlon CE, Han D, Booth M, Motala A, Tariq A, Dudley W, Shanman R, Shekelle PG. Diagnosis of Gout. Comparative Effectiveness Review No. 158. (Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2012-00006-I.) AHRQ Publication No. 15(16)-EHC026-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2016. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

AHRQ Pub. No. 15(16)-EHC026-1-EF
February 2016